THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 6, July-August, 1971
Printed in U.S.A.

Transmitting Television as Clusters of Frame-to-Frame Differences

By J. C. CANDY, MRS. M. A. FRANKE, B. G. HASKELL, and F. W. MOUNTS

(Manuscript received February 16, 1971)

A number of redundancy reduction techniques are used in a coder that is about eight times more efficient than simple PCM. The coder is capable of transmitting Picturephone® *signals at an average rate of one bit per picture-element (2 Megabits per second). When there is movement in the scene, most transmission time is devoted to the parts of the picture that change significantly. The data are generated irregularly but the data flow is smoothed prior to transmission in a buffer that holds about one frame of data. The redundancy reduction techniques used and the behavior of the coder are discussed both from an intuitive and from a statistical viewpoint.*

The positions of elements that change are signaled by addressing the first element of a run of changes and marking the end of the run with a special code word. The changes of luminance are transmitted as frame-to-frame differences using variable-length code words. When rapid motion makes the buffer more than a quarter full, only differences for every second element are transmitted, the values of the intervening changed elements being set equal to the average of their neighbors. If the buffer continues to fill, the threshold that determines which changes are significant is raised from 4/256 to 7/256 of the maximum signal value. When violent motion causes the buffer to fill completely, replenishment is stopped for about one frame while the buffer empties.

Subsampling and raising the threshold are not objectionable because viewers rarely detect the small impairments introduced in moving images. Observers are critical, however, of small impairments in stationary scenes. Thus, to maintain high quality in stationary areas, the entire picture is forcibly updated every three seconds by transmitting 8-bit luminance values for three lines of every frame.

A record of the coder's behavior is available as a 16-millimeter movie film.

I. INTRODUCTION

The introduction of *Picturephone* service has stimulated a great deal of interest in techniques for transmitting video signals more effectively than the way normally used for broadcast television. Since digital transmission over large networks has been found to be more economical for video signals than analog transmission, digital coding is presently receiving most of the attention. One of the best known improvements on simple PCM is differential coding, which takes advantage of the similarity between consecutive samples along a scanning line. Differential coding is simple and yields a bit-rate savings of around 50 percent depending on the picture quality desired.

This paper describes a method that makes use of the similarity of pictures in successive frames as well as the similarity of adjacent samples. Until recently, using the correlation between frames to improve efficiency required complex and costly equipment; but now, with low-cost memories and integrated circuits, it is economically attractive and seems to be particularly suitable for the visual telephone application.

All techniques that reduce the required channel capacity by anticipating a redundancy in the signal have a limitation, in that signals which do not contain the expected redundancy cannot be transmitted unless some alternative process is provided for them. This limitation is possibly the reason why redundancy-reducing techniques have not been used extensively for broadcast television, where there is need to display unusual scenes to attract viewers and entertain them. With visual telephone, however, there is a great need for economical transmission because each signal will not have an audience of millions to share its costs. Indeed it is unlikely that users will want to pay for transmission capabilities that are rarely needed.

Recent work[1–5] has demonstrated how differential coding enables *Picturephone* signals to be transmitted using three or four bits per picture-element. This is about three bits less than simple PCM requires to give comparable quality. Inherent in this technique is a restriction on the amount of detail in the scene that can be reproduced faithfully. A much greater saving has been obtained by F. W. Mounts,[6] using coders that signal only the changes between successive frames. These coders require only one bit of channel capacity per picture-element,

the restriction in this case being on the amount of movement that can be accepted in the scene.

It is anticipated that users of *Picturephone* service will want to display scenes containing fine detail much more often than scenes containing rapid motion; and so they would tolerate blurring of moving objects more readily than a similar loss of spatial detail in stationary scenes.[7] Consequently, ways are used for exploiting the correlation between successive frames as well as that between successive elements in order to effectively encode *Picturephone* video signals.

The work to be described here, which is a continuation of F. W. Mounts' original work, improves the coder so that more motion can be tolerated while using a much smaller buffer than was originally required. (A coder that uses no buffer but requires a data rate of 1.5 bits per element is described in Ref. 8). The techniques used here rely primarily on two phenomena:

(*i*) Large areas of the average *Picturephone* scene change very little or not at all between successive frames. Thus, in each frame only a small amount of new information is required to specify these areas to the receiver.

(*ii*) More information is needed for specifying areas of the picture that change. But they need not be reproduced at the receiver with as much spatial resolution as stationary areas because the storage properties of the camera tube tend to blur movements, and viewers cannot accurately resolve fine details that are in motion.

The coder transmits fresh information describing the stationary parts of the picture at least once in three seconds. At the receiver, this information is retained in a frame memory from which the display is derived. Portions of the picture that change significantly are updated as soon as they are detected, but sometimes with slightly lower resolution than in the stationary areas.

There follows a description of the proposed coding strategy and an account of a simulation of the coder that has been demonstrated.*

II. CONDITIONAL REPLENISHMENT

In Ref. 6 F. W. Mounts called his method "Conditional Replenishment." He transmitted only the information necessary to describe the

* A demonstration of the system processing a live scene was shown during the Keynote Session of the 1970 IEEE International Convention.

intensity of elements that changed between frames. The implementation used two frame memories, one at the receiver and another at the transmitter, to store the required reference picture. The incoming signal was compared element by element with the reference picture. When the magnitude of the frame-to-frame difference was greater than a certain threshold, it was regarded as significant,[9] and the new signal value replaced the reference value in the memory. The new value was also transmitted to the receiver, where it updated the contents of the receiver frame memory. In this way the receiver memory was made to track the transmitter memory, thereby providing a signal for display.

Every time an element value changed, two numbers were transmitted, one describing the new amplitude and the other describing the location of the element. In addition, synchronizing words were sent at the start of every frame and every line, making it necessary to address only the horizontal position of changed elements.

For most television pictures, the technique of sending only data relating to elements that change gives a considerable saving of transmission cost. There is a saving when a large fraction of the scene is stationary or when moving objects contain large areas that are uniformly bright; in either case little transmission is required. Part of the saving is offset by the need to define the position of changed elements in addition to their new values. Nevertheless, F. W. Mounts found that, by using a large buffer to smooth the highly irregular data flow, visual telephone signals could be transmitted using, on the average, only one bit per picture-element.

III. THE PICTURE FORMAT AND TYPICAL STATISTICS

The video signal, which is very similar to the *Picturephone* video signal, is derived from a picture scanned with 271 interlaced lines at 30 frames per second. An example is shown in Fig. 1. The bandwidth is nominally 1 MHz, and elements are sampled at about 2 MHz to give 248 samples in a line period. The visible portion is about 13 cm high and about 14 cm wide; it contains 255 lines each with 207 elements. For practical convenience the signal is coded as 8-bit PCM so that all processing can be performed digitally in real time. The positions of elements in the frame are described by signaling the start of every line and addressing their positions only within the line. If an 8-bit address word is used, 206 of its 256 values are needed to address the visible elements on a scanning line. In the simplest case, only three

Fig. 1—A typical picture.

others are needed: one to mark the start of an even field, one to mark an odd field, and one to mark the start of a new line. In an operational system, however, all of the remaining 50 values would probably be used to help protect the synchronization from error and to signal special modes.

The frame contains a total of 67,208 elements of which 52,785 are visible. Usually only a small fraction of these elements change between any two frames.[10,11] The ordinate in Fig. 2 shows the number of elements that change in a frame-time by more than 3/256 of the maximum signal amplitude. This graph shows 18 seconds of signal measured when there was an unusually large variety of motion in the scene. For convenience the scale is classified into five ranges corresponding to the viewer's subjective judgment of the activity. Figure 3 is a probability density function of the number of changes in a frame. It shows how often the various kinds of activity can be expected in typical *Picturephone* scenes. The data were collected from 75 minutes of picture material which included head-and-shoulder views of one and occasionally two persons, camera panning, subjects walking through the field of view, and pictures of printed text. On the average, less than 4 percent of the elements change in a frame-time.

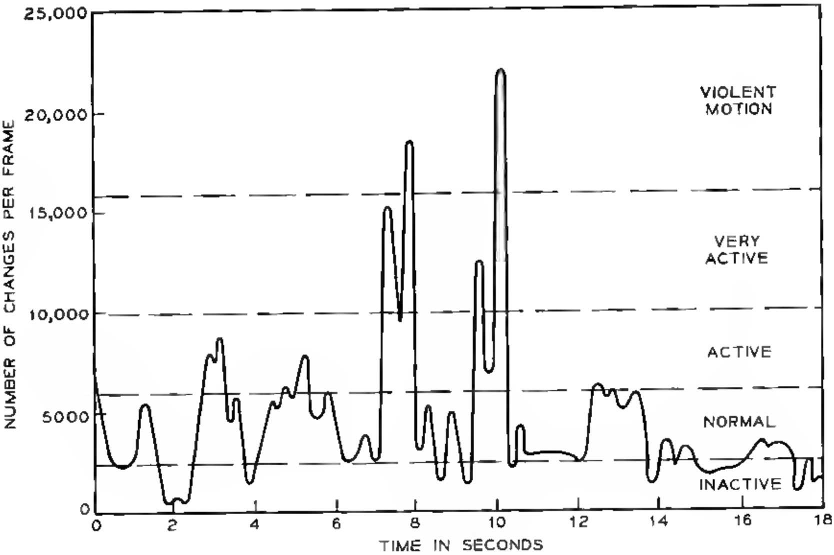

Fig. 2—The number of elements in a frame that change by more than 3/256 of maximum amplitude, plotted over a period of 18 seconds when there was a variety of motion. Each frame contains 67,000 elements of which 53,000 are visible.

IV. THE BUFFER REQUIREMENT

When only 4 percent of the elements change, it is fairly easy to design a coder whose data-rate is less than one bit per element; there are times, however, when the rate is much higher. Thus, a buffer must be used at the transmitter to smooth the highly irregular flow of data from the coder and feed it to the channel at a constant rate. Another buffer at the receiver performs the inverse operation.

It is evident in Fig. 2 that peaks of high activity last for an appreciable part of a second; thus, a buffer to smooth them out would introduce an intolerable delay into the signal path. To prevent delay and echoes in the accompanying voice channel from being a distraction to users of *Picturephone* service, the total delay between talker and viewer should be less than a third of a second;[12] i.e., ten frames. This sets an upper limit on the size of the buffer.

Not much is gained, however, by using a buffer this large. The buffer should be at least large enough to smooth the data over one field-time (two interlaced fields per frame are used throughout), since the changing elements are usually very unevenly distributed within a

picture. This is illustrated in Fig. 4 by bright dots placed in the position of elements that have changed in the picture. There seems to be little advantage, however, in carrying data over from one field to another; unless it can be accumulated for many frames, the activity is usually very similar in successive fields. To illustrate this similarity, Fig. 5 gives the probability that the variation in the number of significant changes in successive fields exceeds a stated amount: the number of changes in a field differs from the number in the previous field by more than 500 in only 5 percent of the fields measured. Another study[13] has also confirmed that the buffer should indeed be capable of storing data over a field-time and not very much larger.

It is proposed that the buffer be capable of storing the amount of data that is transmitted in two field-times. This is more than enough to smooth the irregularities caused by spatial distribution of activity and also allow latitude for controlling the coder. Use of this buffer, which is about one tenth the size of the one used by F. W. Mounts,[6] requires that the data fed into it in a frame-time be approximately

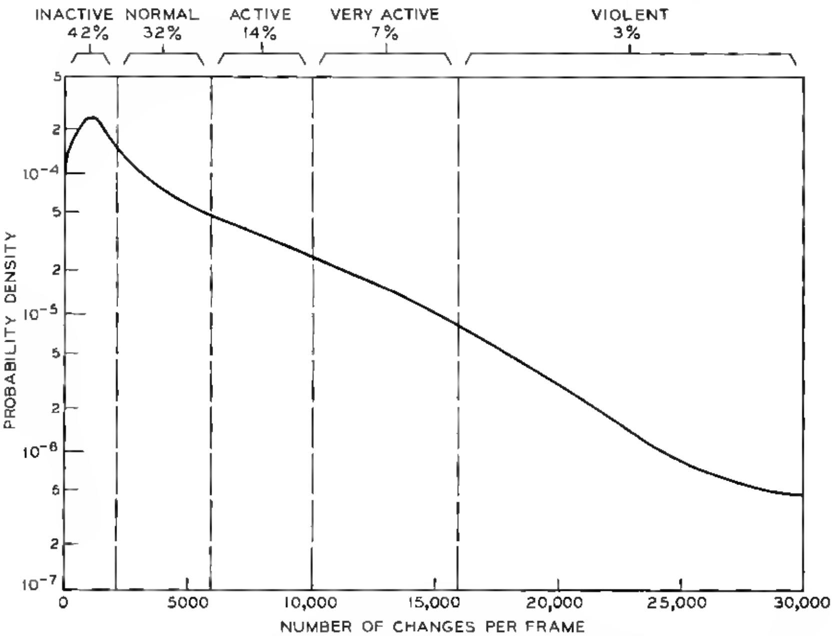

Fig. 3—The probability density function of the number of changes in a frame for 75 minutes of signal. The degrees of motion in the scene are indicated.

Fig. 4—A display of dots marking elements that have changed in a frame during very active motion: about 12,000 changes occurred in this frame. The picture is that shown in Fig. 1.

equal to what can be transmitted by the channel in a frame-time. To satisfy this constraint on the coder, the average number of bits used to signal a change in an element value must decrease as the activity in the scene increases. For example, Table I gives an estimate of the number of bits available to code each change for various amounts of activity if the transmission rate is 67,000 bits per frame (one bit per element for *Picturephone* use):

From Fig. 3 and Table I, it is seen that conditional replenishment accommodates pictures of normal activity, i.e., 70 percent of the scenes, when transmitting two 8-bit words (address and amplitude) for every change. It will be shown how to extend the range by encoding with fewer bits. Elements will be addressed in clusters instead of individually, and their changed amplitudes will be transmitted as frame-to-frame differences using four bits on the average instead of the eight bits needed to define a completely new amplitude.

V. ADDRESSING ELEMENTS IN CLUSTERS

In the original conditional replenishment system, about half of the transmitted data were used for addressing the positions of changed

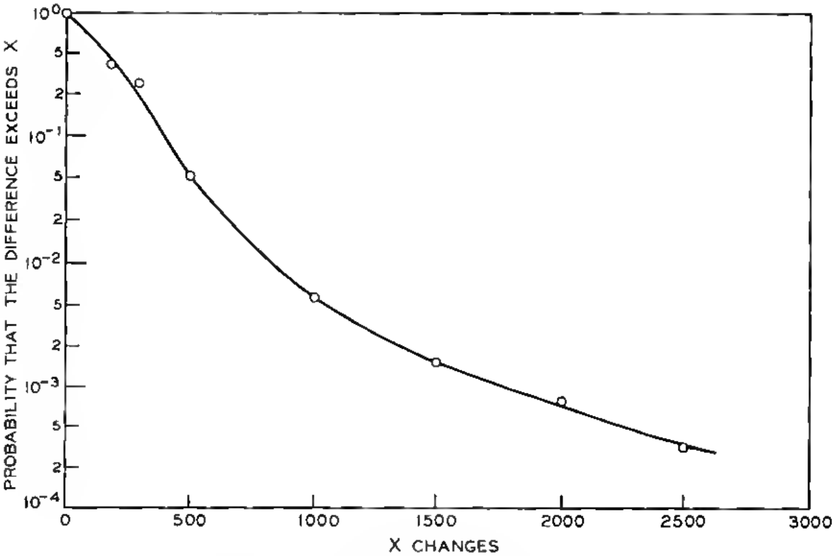

Fig. 5—The probability that the difference between the number of changes in one field and those in the next exceeds X.

elements. It is observed in Fig. 4 that significant changes usually occur in clusters crowded near brightness edges of moving objects. Thus, it is profitable to describe the position of changed elements in groups rather than individually. To explore such methods, the magnitudes of frame-to-frame differences for one whole frame were stored in a memory during a period of active movement. Figure 6 shows an example of how changes are distributed among runs of various lengths. In this example 10,547 elements changed requiring 84,376 bits to address them individually each with eight bits. Using cluster coding, the start of each run is addressed with an 8-bit word, and the end of

TABLE I—NUMBER OF BITS PER CHANGE AVAILABLE AT VARIOUS DEGREES OF ACTIVITY

Description of the Activity	No. of Changes in a Frame of 67,000 Elements	Average Number of Bits Available per Change
Inactive	$<$ 2,000	$>$30
Normal	$\simeq$ 4,000	$\simeq$16
Active	$\simeq$ 8,000	$\simeq$ 8
Very active	$\simeq$13,000	$\simeq$ 5
Violent	$>$16,000	$<$ 4

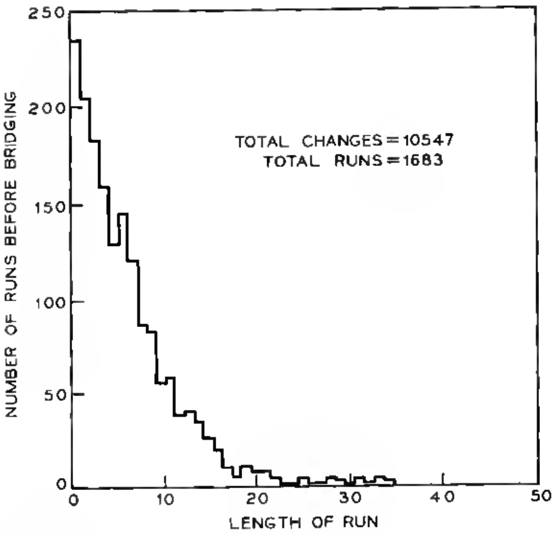

Fig. 6—A histogram of the number of contiguous runs of various lengths in a frame. The motion was very active.

the run is signaled with a special 4-bit word. With this method the 1683 runs of changed elements in Fig. 6 could be positioned using only 20,196 bits; a saving of 64,180 bits, (about 75 percent). It will be shown that an additional saving is possible.

Addressing runs is inefficient when applied to isolated changes in the picture, because 12 bits are used to fix their positions. However, it was observed that most isolated changes can be ignored without spoiling the picture quality. The following criterion provided acceptable quality: *any change in an element, no matter how large, was regarded as insignificant if it were immediately preceded and followed by two elements that changed by less than the threshold of significance.* For the data in Fig. 6, 137 changes could be ignored under this criterion.

Another improvement is obtained by coalescing runs that are separated by a small number of unchanged elements. For example, when using the cluster-coding just described and 4-bit words to signal changes of amplitude, it is preferable to continue a run that is interrupted by less than four unchanged elements, since it requires 12 bits to end a run and restart a new one, but only 4, 8, or 12 bits to continue coding the insignificant changes between runs. The technique of

coalescing runs and ignoring isolated changes will be referred to as "bridging." Bridging often improves quality as well as efficiency because some elements, changing less than the threshold, are updated and isolated noise impulses are suppressed. It should be noted that the suppression of isolated changes takes place before bridging. Performing the operations in the reverse order would be somewhat less efficient.

Figure 7 shows the bright dots covering elements that would be coded using the techniques described above and Fig. 8 shows dots at the start of every bridged-run (cluster). Figure 9 shows the distribution of the cluster lengths for the data in Fig. 6. (Notice how bridging significantly lengthens the runs, reducing the number of runs from 1,683 to 736 while increasing the number of elements in the runs only from 10,410 to 11,886.) To transmit these data, about 47,544 bits would signal 4-bit amplitude changes and 8,832 bits would signal addresses. Figure 10 is plotted the same as Figs. 6 and 9 but with the threshold raised to 7 parts out of 256. Here the changed elements are grouped more closely together, and there are fewer of them.

Fig. 7—A display of elements that would be transmitted using cluster coding for the changes in Fig. 4. Isolated changes are ignored and gaps of three or less are bridged.

Fig. 8—The dots show the start of every cluster.

This frame requires 32,516 bits for amplitudes and 9,132 bits for addresses.

In general, cluster coding reduces the number of address-bits significantly. Application of cluster coding to the data used in Fig. 3 is illustrated in Fig. 11. This graph shows that the number of clusters increases much slower than does the number of changes in the picture. Indeed when there are more than 10,000 changes, the number of clusters is relatively constant; therefore, cluster coding is very effective for smoothing the generation of address bits. The next two sections describe methods used to reduce and regulate the generation of bits needed to signal amplitude-changes.

VI. FRAME-TO-FRAME DIFFERENCES

6.1 *Amplitude Distribution of Difference Signals*

While discussing cluster coding, we allowed 4-bit words for describing frame-to-frame differences of the video signal. This difference signal is generated at the transmitter when the reference is subtracted from the input. Its amplitude can range between ±255 units but usually it is small, as Fig. 12 demonstrates. Figure 12 shows how the

magnitudes of the significant frame differences are distributed statistically for typical *Picturephone* scenes. The motion in the scene was normal for curve (a) and very active for (b). The circuit used for generating the frame differences is shown in Fig. 13. The quantizer was bypassed while the data was being taken.

The frame-to-frame difference signal may be quantized and transmitted with much less than the eight bits needed to send the absolute amplitudes. For example, when the quantizer is used in the circuit of Fig. 13 (output levels correspond to the eight divisions of the abscissa of Fig. 12), the displayed picture has good quality for still and slowly moving scenes. However, there is noticeable distortion of rapidly moving edges: a sharp, moving edge appears as a cascade of smaller steps as each increase of luminance corresponds to the magnitude of the outer quantization level (43 units). This distortion is eliminated by providing more quantization levels, for example 64 levels can give excellent reproduction for all types of activity in the scene.

Another important reason for using more than 16 levels is to reduce

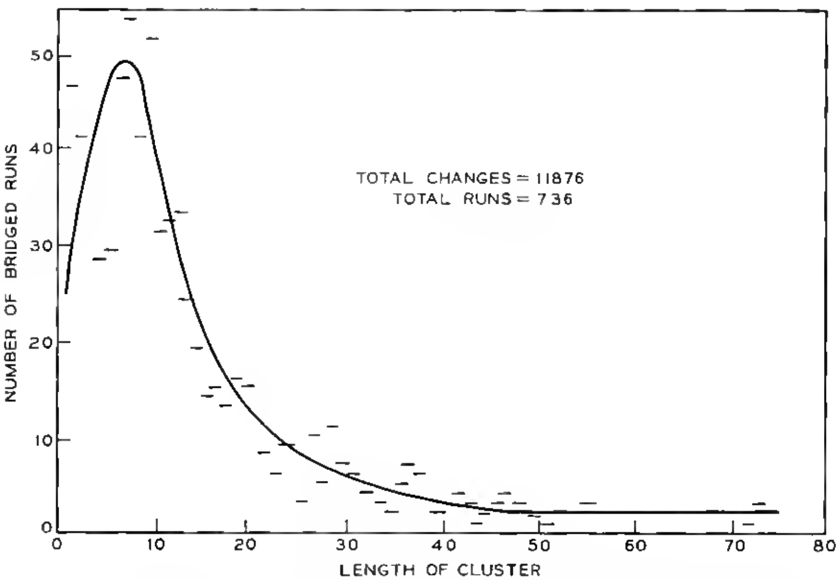

Fig. 9—A histogram of the number of clusters of various lengths for the runs shown in Fig. 6. Isolated changes are ignored and gaps of three or less are bridged.

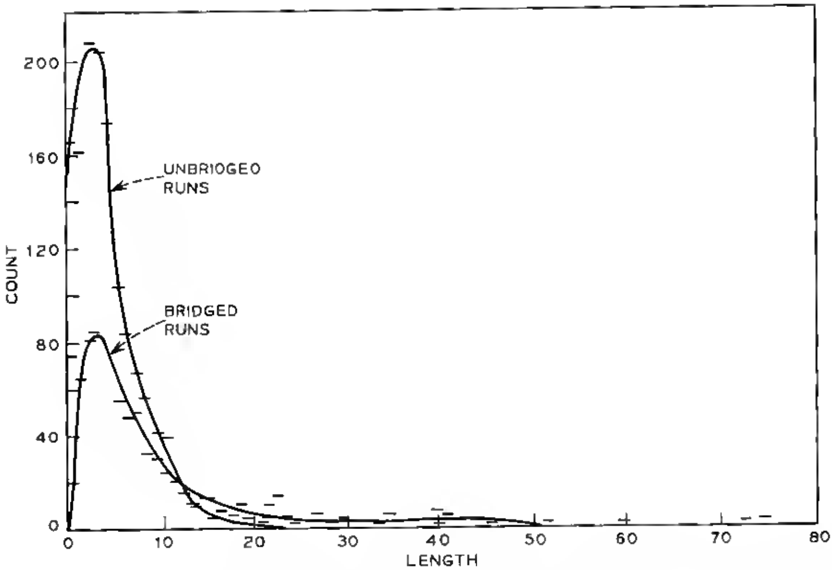

Fig. 10—The effect of raising the threshold from 4/256 to 7/256 on the runs and bridged-runs given in Figs. 7 and 9. Changes that equal or exceed the threshold are counted.

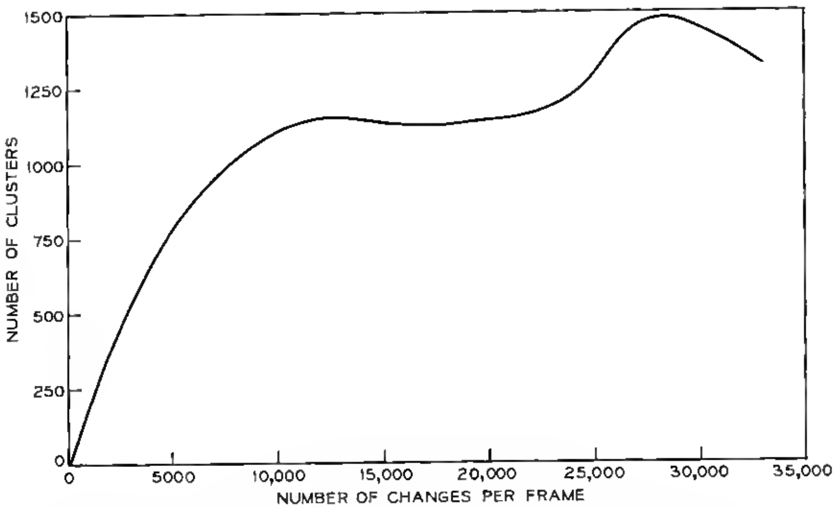

Fig. 11—The average number of clusters in a frame plotted against number of changes. The threshold was 4/256.

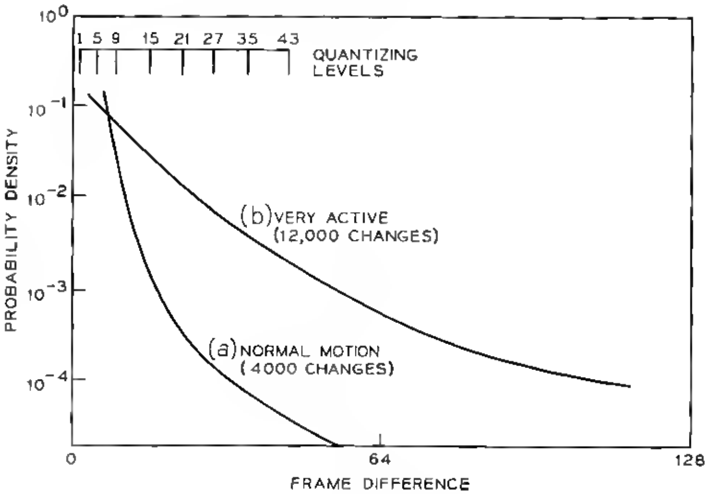

Fig. 12—The probability distribution for the amplitude of frame-to-frame difference signals during normal and active motion. Its maximum is 255 units.

the residual discrepancy between the reference and the input. The stored reference is updated by addition of the quantized difference signal; therefore, its updated value will differ from the input by an amount dependent upon the quantizing scale even after movement has ceased. If, in the next frame time the discrepancy exceeds the threshold, transmission capacity will be used to correct it even though the input signal remains unchanged. This need for additional transmission, after large changes have ceased, makes the system inefficient for certain types of input. Therefore, the coder uses a quantizer that can represent any change to within ± 4 units of its true value. The quantizing scale

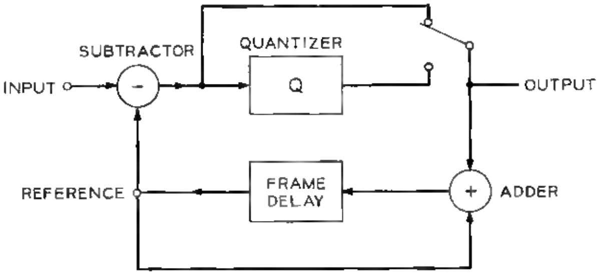

Fig. 13—The circuit used for measuring frame-to-frame difference.

has 64 levels distributed thusly:

$$\pm 1,\ \pm 5,\ \pm 10,\ \pm 15,\ \pm 20,\ \pm 27,\ \pm 35, \cdots$$

increasing in increments of ± 8 up to ± 235.

A signal quantized to 64 levels can be signaled with 6-bit words, but since the outer levels are rarely used, we use only a 4-bit word to signal the innermost 14 levels. The 4-bit word has sixteen distinct values. One of the remaining two values is used to specify the end of a cluster. Whenever the magnitude of the next frame difference exceeds 39 units the remaining 4-bit word is used to tell the receiver that the next 12 bits should be decoded as two 6-bit words rather than three 4-bit words. These 6-bit words specify the frame differences using the full set of 64 levels. Six-bit words continue to be transmitted in pairs until the last word represents a level lying in the innermost 14 levels; then we revert to 4-bit words.

Figure 14 shows a picture with bright spots marking the elements that are updated with 6-bit words; they occur in groups near large rapidly moving edges in the picture. On the average, 6-bit words are used for less than 10 percent of the updated elements. We find that

Fig. 14—The changes that are signaled with 6-bit words for the data in Fig. 4.

this method of transmitting quantized frame differences requires only about half the transmission capacity needed for sending the true picture-element brightnesses as 8-bit PCM. However, when differences are transmitted, a mechanism is needed for preventing transmission faults from introducing permanent errors into the signal. This mechanism is described in section 8.2, 'Forced Updating.'

6.2 *Subsampling*

The techniques described in the previous sections allow pictures to be transmitted using, on the average, less than six bits per changed element. Thus, pictures with active motion can be accommodated. However, to be useful, the range of the system must be extended even further. This extension is obtained by making use of the fact that in moving parts of the picture the sampling frequency can be halved, yet not cause noticeable degradation.[7]

A measure of the activity in the scene is obtained by monitoring the number of bits held in the buffer. When it exceeds a prescribed amount, indicating active motion, we switch to a subsampling mode of operation. In this mode only the changes in every other element in a cluster are transmitted and used to update the stored reference pictures. The amplitudes of the intervening elements in the cluster are set equal to the average of the two neighboring values. Table II shows the states of a sequence of sixteen elements having, initially, reference values A_1, A_2, $\cdots$, A_{16}. The new input values are assumed to be those listed on line (b) where, in every case, the change from value A to value B exceeds the threshold. Line (c) shows which difference signals are transmitted in the normal mode of operation. The change A_3 to B_3, an isolated change, is not transmitted, but the stationary element value A_{10} is transmitted in order to bridge the cluster. Line (d) shows the new reference, the apostrophe signifying that a quantization has occurred in the value. If the coder were operating in the subsampling mode, the transmitted differences would be those marked on line (e), and the new reference would be those values on line (f). Here, the symbol C represents the average of values on either side of the element; for example,

$$C_9 = \frac{B_8' + B_{10}'}{2}$$

When subsampling, the set of elements whose values are considered for transmission are arranged in a fixed checkerboard pattern in the raster. This pattern is unchanged from frame to frame in order that discrepancies between the reference and the input caused by the inter-

TABLE II—SELECTION OF ELEMENTS FOR TRANSMISSION IN BOTH THE NORMAL AND THE SUBSAMPLE MODE

A_n represents the original reference amplitude of the nth element
B_n represents a new value of an amplitude
B_n' represents a quantized amplitude
C_n represents an interpolated amplitude

	(a) Reference	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{16}
	(b) Input	A_1	A_2	B_3	A_4	A_5	A_6	B_7	B_8	B_9	A_{10}	B_{11}	B_{12}	A_{13}	A_{14}	A_{15}	A_{16}
Normal Mode	(c) Elements selected for transmission							X	X	X	X	X	X				
Normal Mode	(d) New reference	A_1	A_2	A_3	A_4	A_5	A_6	B_7'	B_8'	B_9'	B_{10}'	B_{11}'	B_{12}'	A_{13}	A_{14}	A_{15}	A_{16}
Subsample Mode	(e) Elements selected for transmission								X		X		X				
Subsample Mode	(f) New reference	A_1	A_2	A_3	A_4	A_5	A_6	C_7	B_8'	C_9	B_{10}'	C_{11}	B_{12}'	C_{13}	A_{14}	A_{15}	A_{16}

polation shall not instigate a transmission; moreover, a changing pattern is subjectively unpleasant.

Using the subsampling technique, very active motion can be accommodated in the scene, because, on the average, less than four bits are needed to signal each significant change. Although the spatial resolution is effectively reduced by a factor of two in regions of motion, it is almost impossible to see. Indeed, additional loss of resolution can be tolerated to further extend the range of the coder. This extension is obtained by raising the threshold[6] that determines which changes are significant.

6.3 *The Threshold*

Normally the threshold is set to accept changes greater than or equal to 4 parts out of 256; this gives excellent reproduction of motion, yet prevents small amounts of noise from needlessly using transmission capacity. When there is active motion in the picture and the buffer starts filling in spite of subsampling, the threshold is raised one unit at a time as the buffer fills up, finally reaching a maximum threshold of seven units.

When the threshold is at a high value the unreplenished changes appear as a stationary noise, which has the appearance of a dirty windowpane placed before the scene. With threshold values in excess of eight this noise is clearly visible in dark, moving areas of the picture. When the threshold is less than eight the picture impairment is small, and with thresholds less than five the noise is unnoticeable. Figures 15 and 16 show how raising the threshold reduces the number of changes that are counted as significant. Besides reducing the transmission requirement, raising the threshold also makes the coder more tolerant of noise in the input video. In fact, the coder is capable of processing signals that have been contaminated with noise whose root-mean-square value is 35 dB less than the peak signal value. Such high noise levels would be unacceptable for commercial service; we expect noise to be 45 dB less than the signal.

VII. BUFFER OVERLOAD

The techniques we have described so far allow the majority of *Picturephone* scenes to be encoded and signaled at an average rate of one bit per picture-element. Only occasionally does the motion become so violent that the system is congested. Such situations occur when the camera is panned over a very detailed scene or when the

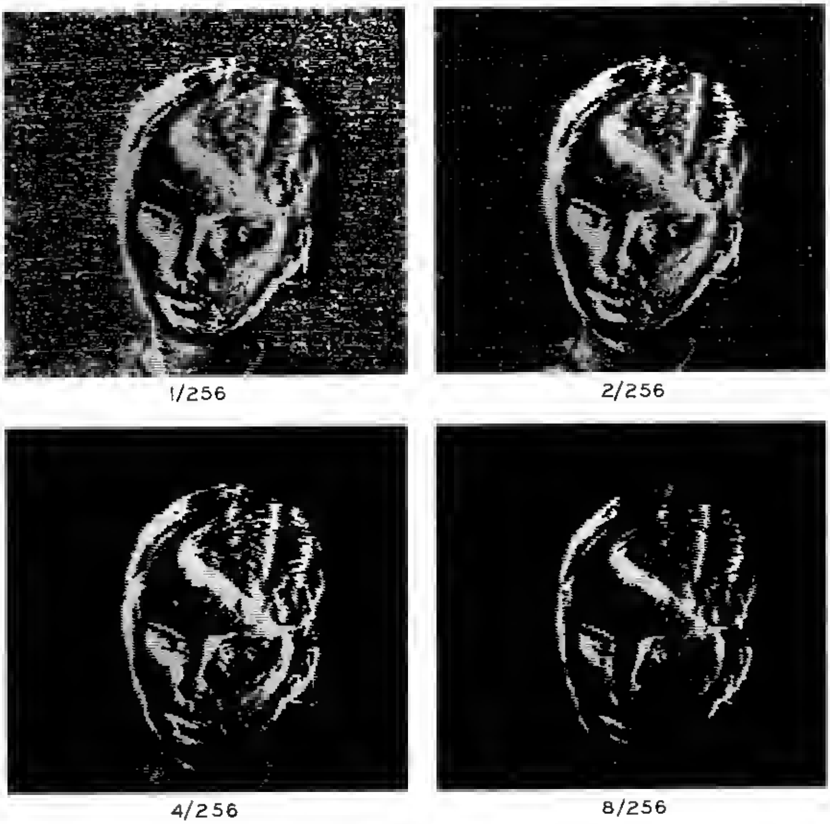

Fig. 15—A display of the elements whose changes equal or exceed thresholds of 1/256, 2/256, 4/256, 8/256.

subject suddenly stands up close to the camera; these instances are usually of short duration. To prevent the buffer from overflowing at these times, we simply inhibit all replenishment of memories at the end of the cluster during which the buffer becomes 99 percent full. Replenishment is inhibited until the buffer content falls to about 2,500 bits, which takes about one frame-time. Afterward, coding resumes with the constraint that it remain in the subsampling mode with a threshold of seven for at least one entire field.

While replenishment is inhibited, the data previously stored in the memory are displayed again, and the motion in the picture becomes somewhat jerky.[14] Viewers have not found the operation to be very

objectionable, and some people have grown accustomed to seeing jerkiness in vigorous movement. However, this is one aspect of the coder that can probably be improved by using the vertical correlations in the picture[8] in much the same way as subsampling uses the horizonal correlations.

VIII. FIXED BACKGROUND TRANSMISSION

8.1 *Start-of-Line Codes*

In previous sections we have concerned ourselves with describing techniques for signaling changes which occur irregularly in the picture. Signaling addresses and amplitude changes use about 90 percent of the transmission capacity; the remaining 10 percent represents the synchronization and forced updating. These data are transmitted independently of picture content unless the buffer fills; then the forced updating is interrupted for a frame-time but the synchronization is always transmitted in order to keep the transmitter and receiver in step.

We have already mentioned that the start of each new line of the raster is signaled with a selected value of the 8-bit address words. (A

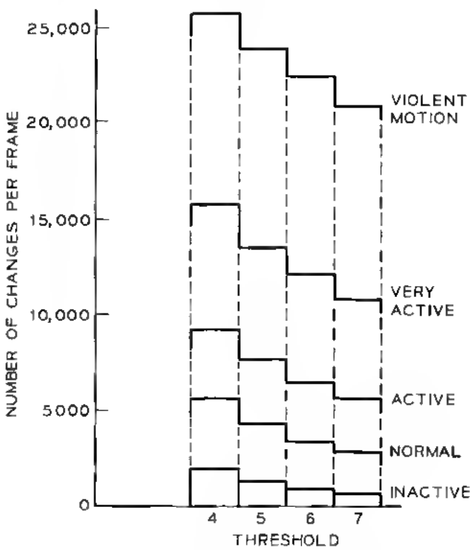

Fig. 16—A histogram of the number of changes in a frame that exceeds various thresholds for typical scenes.

suitable choice of the code will not be discussed in this paper, but notice that 50 code words are available after 206 have been reserved for addressing the positions of elements on the line.) Changes from the normal to the subsampling mode will be signaled by a change in the start-of-line code. A special code is also used at the start of the first line in each field and at the start of lines that are to be forcibly updated.

Using 8-bit words for marking starts of lines requires about 2,000 bits in each frame-time, which is about 3 percent of the intended transmission capacity.

8.2 *Forced Updating*

Forced updating is a process for periodically rewriting the entire picture with full amplitude values. It serves two purposes: To insure a very high quality reproduction of stationary parts of a scene and to correct errors introduced during transmission by periodically aligning the contents of the transmitting and receiving memories.

This forced updating is accomplished by transmitting in each frame-time, three lines of the picture as 8-bit PCM. The three lines are evenly spaced in the raster and are moved up each frame-time so that the entire picture is updated in about three seconds: Thus, less than three seconds after a portion of the picture became stationary it acquires the quality obtainable with 8-bit-PCM transmission. Nearly perfect reproduction can be obtained for graphics and similar scenes that display spatial detail rather than movement.

Forced updating of three lines in each frame uses about 5,000 bits or about 8 percent of the transmission capacity. Thus, about 11 percent of the transmitted information is fixed.

IX. PREVENTING THE BUFFER FROM EMPTYING

We have described ways for preventing buffer overflow by reducing the rate at which data are generated, always taking care to maintain a satisfactory quality in the transmitted picture. The problem of preventing the buffer from becoming empty is much easier to solve because there are many ways of acquiring data in order to improve the quality of the picture. We have chosen to use the forced updating mechanism for this purpose. The next line of input data is "forcibly updated" whenever the buffer content falls below 2,500 bits. When there is no motion, the entire picture is updated, within a quarter of a second, by this technique. This helps to obtain a high-quality repro-

duction of stationary scenes, such as graphic material, and always provides data for the buffer even during the vertical fly-back interval.

X. THE SIMULATION TRANSMISSION EXPERIMENT

Having described the coding strategy, we will now describe a laboratory simulation of the system.

The simulator used produces a received picture which has all the characteristics of one processed by a real system except that effects of transmission delay and digital transmission error are not included.

The simulator is a more complex version of the test circuit shown in Fig. 13; its main features are given in Fig. 17. The video signal enters at the upper left where it is sampled and coded as 8-bit PCM. This digital coding is required because it is impractical to build analog memories and analog processing circuits that are sufficiently accurate to perform the operations previously described. The coded input signal is compared with the reference signal emerging from the frame memory. The difference is then fed to a quantizer and to a threshold circuit that determines whether or not the change is significant. If the difference is insignificant, the threshold circuit tells the control

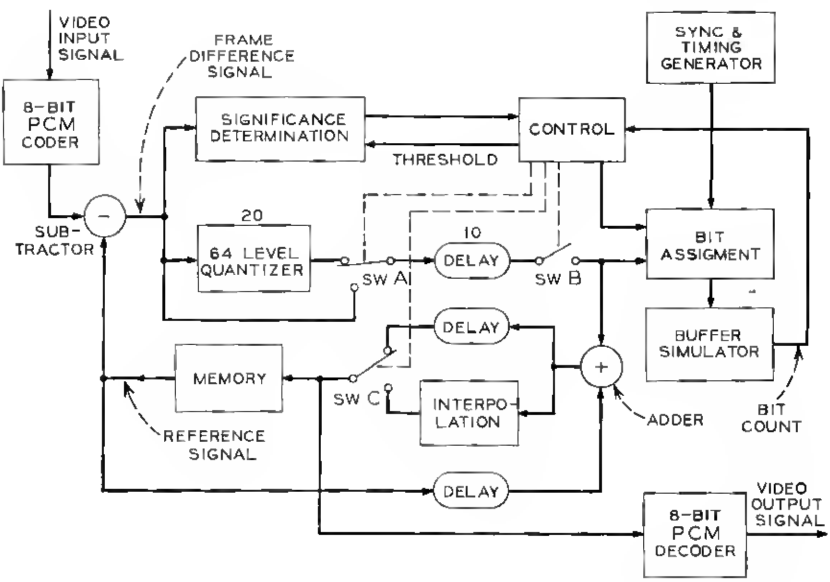

Fig. 17—The laboratory system used for simulating the behavior of the coder. It operates directly on real video signals.

to open switch B to prevent the quantized signal from being fed to the buffer or to the memory. When the first significant difference is encountered, starting a cluster, the control closes switch B so that the quantized difference can be fed toward the buffer and to the adder. The adder combines the difference with the reference that is being circulated from the output of the frame memory. Thus, for significant changes the reference picture is brought up-to-date by adding the quantized difference to it. For insignificant changes, it remains unchanged in the memory since zero is added to it.

The purpose of the delay included before switch B is to give the control circuit enough time to decide whether to start or end a cluster. Recall that changes not accompanied by other changes are ignored, and clusters are ended only when the next four picture-elements change less than the prescribed threshold. The delay in the memory feedback path insures that the reference is added to the corresponding quantized difference.

In a real system the significant frame-differences would be coded and transmitted to the receiver to update the contents of its memory as is done at the transmitter. But in the simulator we avoid the need for a transmission link and a receiver by taking the display signal from the input of the frame memory. Although we have avoided the need for transmission, we still need an indication of the fullness of the transmitting buffer in order to control the behavior of the coder as the buffer fills up. This is accomplished by counting the number of bits that would be sent to a buffer in a real system, and subtracting the amount that would be transmitted. The circuit labeled "bit assignment" determines the number of bits that should be fed to the counter in order to simulate a buffer. It receives the quantized differences, and determines the number of bits needed to code them. It also receives signals from the control circuit which mark the start and end of each cluster of changing elements and a signal to indicate when forced updating occurs. From the address generator it receives signals that mark the start of each scanning line.

A count of the number of bits in the buffer is fed to the control circuit where it is used to determine the operating mode of the coder. The diagram in Fig. 18 illustrates this control function; it shows that the threshold is raised from four to five, and then to six, and finally to seven as the count increases from 10,000 to 20,000, to 35,000, and to 50,000. At a count of 65,000 all replenishment is stopped until the count falls below 2,500. The system starts subsampling when the count exceeds 20,000 and does not return to full sampling until the count

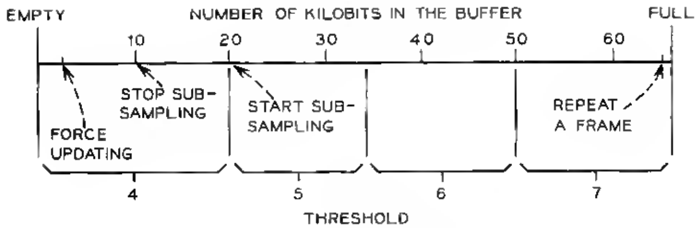

Fig. 18—A representation of the control functions associated with the buffer.

falls to 10,000.* The subsampling is accomplished by simultaneously opening switch B and moving switch C to its lower position for every other element in the cluster. The signal fed to the memory and to the output is then an interpolation of adjacent values.

Switch A is closed to bypass the quantizer and switch C removes the interpolation when lines of the picture are "force updated" with their true values.

XI. BEHAVIOR OF THE CODER

11.1 *Qualitative Behavior*

When the simulator processes the signal in a *Picturephone* transmission, very little impairment can be seen in the received picture. Indeed the received picture appears to be the same as the transmitted one except for two conditions which rarely occur for normal scenes. One is the jerkiness introduced into scenes that change violently over a large area; the other is a moiré pattern that can be seen when a graticule of vertical bars moves enough to require the subsampling mode. The moiré pattern, or aliasing, is a transient that vanishes as soon as the graticule becomes stationary and full sampling resumes.

11.2 *Quantitative Behavior*

Figure 19a shows the probability of buffer content exceeding Q bits with a subject moving vigorously. Figure 19b shows the corresponding probability density function. We see that, most of the time, the buffer is almost empty; only two percent of the time is it more than three quarters full. This indicates that there is considerable advantage in sharing buffers and transmission channels amongst several coders. Figure 20 shows how many bits are used, on the average,

* Choice of these parameters is not critical.

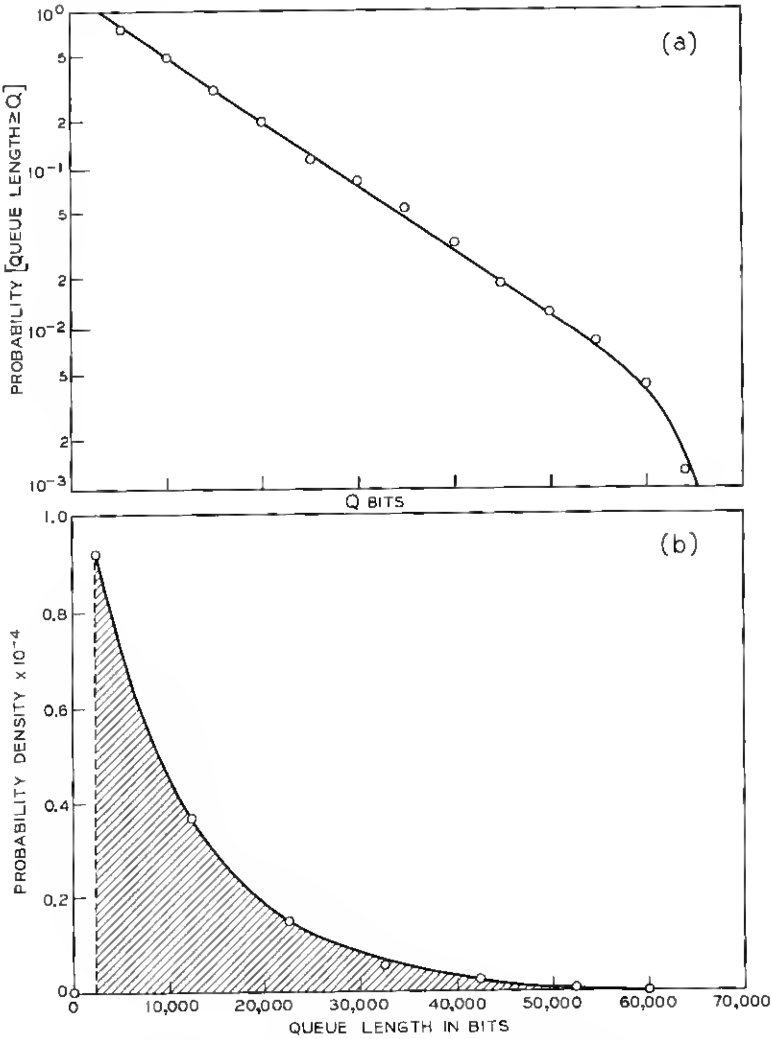

Fig. 19a—The probability of the buffer content exceeding Q bits.

Fig. 19b—The corresponding probability density function.

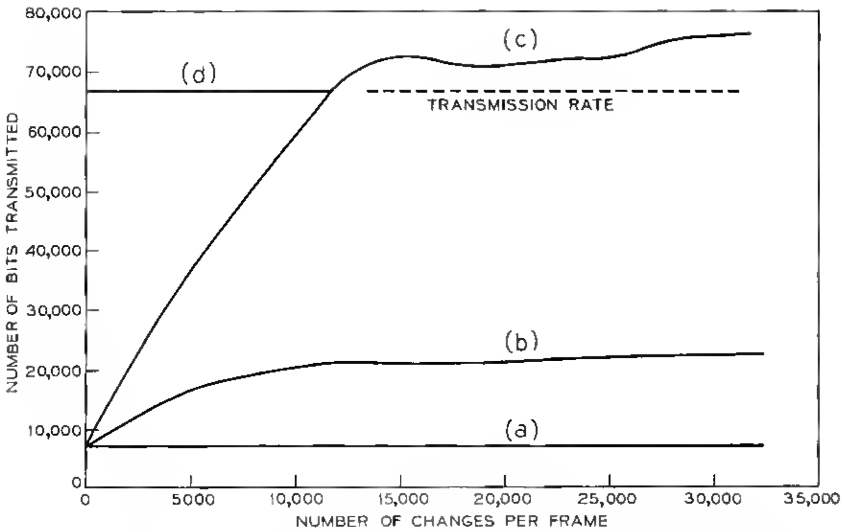

Fig. 20—The average number of bits transmitted plotted against the number of significant changes. The probability of the changes occurring is given in Fig. 2. The ordinate of (a) represents the fixed transmission. The separation (a)-(b) represents the addresses, c.f. Fig. 11; (b)-(c) represents the amplitude difference; and (c)-(d) represents the forced transmission.

for the various functions of the coder at different levels of activity in the scene. The ordinate of the straight line (a) represents the fixed transmission. The distance between curves (a) and (b) represents the bits used for addressing clusters of changed elements, and the distance between (b) and (c) represents the bits used for signaling their changes of amplitude. The change in slope of (c) near 12 kilo-changes per frame and the subsequent lower rate of rise of the curve is caused by the subsampling mechanism. The distance between curves (c) and (d) represents the bits that are generated by forcibly updating lines of the picture in order to prevent the buffer from emptying. On this graph we see that the data generated per frame is less than the transmission rate, provided no more than 12,000 elements change in a frame time; i.e., 90 percent of the time. When more than 12,000 elements change, the extra bits accumulate in the buffer. Only when more than 20,000 elements change for several consecutive frames, i.e., for very violent motion covering a large area of the picture, is the overload mode needed.

XII. CONCLUSION

Taking 8-bit PCM as the standard of coding quality, we have described a method of improving transmission efficiency by a factor of nearly eight, permitting pictures to be transmitted at an average rate of one bit per element. Most viewers agree that there is little visible difference between the processed picture and the original. The quality is certainly better than has been obtained using 16-level differential coding.

The improvement in efficiency is largely obtained by avoiding the use of transmission capacity for elements that do not change from frame to frame. Additional saving is obtained by taking advantage of the fact that reduced resolution is tolerable in changing scenes. Thus, frame-to-frame differences are quantized in amplitude, and sampled at half of the Nyquist rate. In this way the number of bits generated in every field is kept approximately equal to that which can be transmitted in a field time. A buffer is used to smooth the data flow over a field-time, its fullness serving as an indication of the rate at which the data is being generated. This measure is used for controlling the behavior of the coder.

Important subjects which have not been described here are the possibilities of using line-to-line correlation in the coder, the advantages of mixing the data from several coders in order to obtain a more even data flow, and the effect of having an input signal that has already been modulated or coded in a prior transmission.

XIII. ACKNOWLEDGMENT

The authors thank C. C. Cutler and L. H. Enloe for direction and encouragement, D. J. Connor, J. O. Limb, and R. F. W. Pease for their cooperation, R. L. Eilenberger for providing the camera system, W. G. Scholes for the frame memories, and W. Kaminski for the master clock generator.

REFERENCES

1. Maunsell, H. I., and Millard, J. B., "Digital Coding of the Video Signal," B.S.T.J., *50*, No. 2 (February 1971), pp. 459–479.
2. Limb, J. O., and Mounts, F. W., "Digital Differential Quantizer for Television," B.S.T.J., *48*, No. 7 (September 1969), pp. 2583–2599.
3. Brown, E. F., "A Sliding-Scale Direct-Feedback PCM Coder for Television," B.S.T.J., *48*, No. 5 (May-June 1969), pp. 1537–1553.
4. Bosworth, R. H., and Candy, J. C., "A Companded One-Bit Coder for Television Transmission," B.S.T.J., *48*, No. 5 (May-June 1969), pp. 1459–1479.

5. Connor, D. J., Pease, R. F. W., and Scholes, W. G., "Television Coding Using Two-Dimensional Spatial Prediction," B.S.T.J., *50*, No. 3 (March 1971), pp. 1049–1061.
6. Mounts, F. W., "A Video Encoding System With Conditional Picture-Element Replenishment," B.S.T.J., 48, No. 7 (September 1969), pp. 2545–2554.
7. Pease, R. F. W., and Limb, J. O., "Exchange of Spatial and Temporal Resolution in Television Coding," B.S.T.J., *50*, No. 1 (January 1971), pp. 191–200.
8. Limb, J. O., and Pease, R. F. W., "A Simple Interframe Coder for Video Telephony," B.S.T.J., this issue, pp. 1877–1888.
9. Teer, K., "Some Investigations on Redundancy and Possible Bandwidth Compression in Television Transmission," Thesis Technische Hogeschool Delft, September 1959.
10. Seyler, A. J., "Probability Distribution of TV Frame Differences," SMIREE Australia Proc., November 1965, p. 355.
11. Pease, R. F. W., "Frame Difference Signals for a Personal Communication Television System of 271 Lines," Private Communication.
12. Brainard, R. C., Cutler, C. C., and Rowlinson, D. E., "*Picturephone* Communication with Delay," Private Communication.
13. Limb, J. O., and Haskell, B. G., "Buffering of Data Generated by the Coding of Moving Images," Private Communication.
14. Mounts, F. W., and Pearson, D. E., "Apparent Increase in Noise Level When Television Pictures are Frame-Repeated," B.S.T.J., *48*, No. 3 (March 1969), pp. 527–539.